

Article

A Clustering-Based Hybrid Support Vector Regression Model to Predict Container Volume at Seaport Sanitary Facilities

Juan Jesús Ruiz-Aguilar ^{1,*} , José Antonio Moscoso-López ¹, Daniel Urda ² ,
Javier González-Enrique ³ and Ignacio Turias ³

¹ Department of Industrial and Civil Engineering, Polytechnic School of Engineering, University of Cadiz, 11202 Algeciras, Spain; joseantonio.moscoso@uca.es

² Grupo de Inteligencia Computacional Aplicada (GICAP), Departamento de Ingeniería Informática, Escuela Politécnica Superior, Universidad de Burgos, Av. Cantabria s/n, 09006 Burgos, Spain; durda@ubu.es

³ Department of Computer Science Engineering, Polytechnic School of Engineering, University of Cadiz, 11202 Algeciras, Spain; javier.gonzalezhenrique@uca.es (J.G.-E.); ignacio.turias@uca.es (I.T.)

* Correspondence: juanjesus.ruiz@uca.es

Received: 7 November 2020; Accepted: 22 November 2020; Published: 24 November 2020



Abstract: An accurate prediction of freight volume at the sanitary facilities of seaports is a key factor to improve planning operations and resource allocation. This study proposes a hybrid approach to forecast container volume at the sanitary facilities of a seaport. The methodology consists of a three-step procedure, combining the strengths of linear and non-linear models and the capability of a clustering technique. First, a self-organizing map (SOM) is used to decompose the time series into smaller clusters easier to predict. Second, a seasonal autoregressive integrated moving averages (SARIMA) model is applied in each cluster in order to obtain predicted values and residuals of each cluster. These values are finally used as inputs of a support vector regression (SVR) model together with the historical data of the cluster. The final prediction result integrates the prediction results of each cluster. The experimental results showed that the proposed model provided accurate prediction results and outperforms the rest of the models tested. The proposed model can be used as an automatic decision-making tool by seaport management due to its capacity to plan resources in advance, avoiding congestion and time delays.

Keywords: maritime transport; container forecasting; support vector regression; self-organizing maps; machine learning; hybrid models

1. Introduction

Over the last decades, ports have played an important role in international trade and most of the overseas shipping of products is aboard deep-sea container vessels [1]. The increase in traffic of goods and the European unification has led to consider the enhancement of security at border crossings of the European Union. In this sense, the Border Inspection Posts (BIPs) were created in order to guarantee the security at border crossings and the quality of the import-export goods. BIPs are the approved facilities where the checks of goods (transported within containers by trucks or towing vehicles) are carried out before entering the Community territory. However, the sustained growth in the worldwide exchange of goods is creating the need for further inspections resulting in congestion and high load-peaks within the sanitary facilities. This causes time delays and higher costs in the supply chain. The BIPs are thereby bottlenecks that must be necessarily assessed by Port Authorities in order to keep the quality level of the port and avoid losing competitiveness. In order to avoid time delays and congestion in the sanitary facilities, the port management should be able to accurately

forecast the number of container passing through these facilities. An accurate prediction of the volume of container traffic through the BIP may become a useful tool to improve human resources, planning operations and the service quality at ports.

Forecasting time series have focused attention of many researchers for a long time. Many efforts have been carried out to improve the existing methodologies and to achieve enhanced models to obtain accurate predictions in any forecasting fields (for instance, energy consumption, transportation, environment, economy, medicine and health). The focus of this study is mainly set on maritime transport due to the traffic associated with the port BIPs. The proposed forecasting techniques can be divided into three categories: single methods, combined methods and hybrid methods.

The first class comprises both linear and nonlinear techniques. Linear techniques are based on the assumption that a linear relationship exists between the future values and the current and past values of the time series. One of the linear models that has attracted more attention over the past few decades is the well-known autoregressive integrated moving averages (ARIMA) model and its further extended versions. Based on the Box and Jenkins methodology [2] the ARIMA models have been constantly applied to solve forecasting tasks related to maritime transport. ARIMA models were successfully applied to predict export operations in container terminals [3], to predict certain traffic flows of goods that pass through a port [4,5] and recently for predicting container throughput volumes at ports [6].

Moreover, nonlinear techniques have become a strength alternative against the weaknesses of linear models. This forecasting technique has proved to be effective when time series show nonlinear patterns, overcoming the main constraints of linear models. Real-world problems are complex and mostly generated by an underlying nonlinear process [7], such could be the case of volume of containers demand at BIPs. In this subcategory, two machine learning techniques highlight: artificial neural networks (ANNs) and support vector machines for regression (SVR). Instead of ANNs, SVR has attracted increasing attention in recent years due to its inherent abilities to overcome some general limitations of ANNs, such as the achievement of a global minimum. Due to its great generalization ability, SVR has been used in forecasting transport tasks with promising results. These two techniques have been constantly compared in the research literature.

The second category comprises the combined models. These models attempt to overcome the constraints of the single forecasting models when time series exhibit seasonal features, temporal evolutions, correlations, and other patterns hard to capture by single models. To this aim, single methods are combined to seize the abilities of each one in certain forecasting tasks. One of the most frequently used approaches consists of combining a single prediction technique (mainly a soft computing technique) with a clustering method. Based on the divide-and-conquer principle, clustering methods allow to divide the original database into a number of smaller groups, called clusters. The main assumption is that simpler clusters are easier to solve than addressing the whole database. When the clustering method has divided the database into several clusters, a prediction technique is then applied in each cluster independently. Self-organizing maps (SOMs) [8] is probably the best-known clustering method. This popular unsupervised learning technique has been proved to be faster and more accurate and efficient than other clustering methods [9], improving the final prediction performance in time series [10]. A combined SOM-ANN model was introduced by Chen et al. [11] to predict traffic flows in transportation. This work also compared the forecasting performance of the proposed model to the obtained with a SOM-ARIMA model and a single ARIMA model. Results showed that the SOM-ANN model outperformed the rest of models. Due to the recent emergence of SVR in transportation, there is hardly any research related to transport combining SOM and SVR in a two-stage procedure. Nevertheless, it is a widespread solution in many other forecasting fields [12].

The third category includes hybrid models. Single linear models have shown to have several constraints, limitations and disadvantages in particular situations and certain applications (for instance, when the underlying generating mechanism is nonlinear or uncertainty is presented) [13]. In a similar manner, the use of nonlinear models can be totally inappropriate if linear patterns are presented. Moreover, real-world time series are not completely linear or nonlinear, but rather contain both

components. Thus, a methodology using linear and nonlinear models in a hybrid way takes the capabilities of both models, improving the forecasting accuracy and reducing the risk of failure when an unsuitable single model is used. Hybridizing linear models and machine learning techniques have been proposed in recent years to forecast transportation time series. Particularly, ARIMA has been the most commonly used linear model in literature to construct this kind of hybrid model. The first works were proposed considering ANNs as the machine learning technique in [7,14] to forecast time series and they concluded that hybrid ARIMA-ANN models were superior to single SARIMA and ANN models. Since then, many research works can be found in the literature in many areas linked to time series analysis [15,16]. Several authors were also proposed a hybridization of SARIMA and SVR to address several forecasting tasks outside the transport sector [17,18], although it is less widespread compared to SARIMA-ANN models. Related to maritime transport, Xie et al. [19] proposed several hybrid approaches in a comparative way including the SARIMA-SVR model for container throughput forecasting. All these previous studies coincided in pointing out that a hybrid strategy considering ARIMA and SVR models overcame the performance of single models in their respective domains.

In this study, a combined-hybrid forecasting model is proposed in such a way that a hybrid model (SARIMA-SVR) is combined with a clustering method (SOM) to forecast the daily number of containers passing through a BIP, thus resulting in a SOM-SARIMA-SVR model. This methodology unifies the strengths of clustering methods in decomposing the forecasting task into some relatively easier subtasks (using a SOM method) and the strengths of hybrid models to fit linear and nonlinear components (using a SARIMA-SVR model). Thus, the final aim of this study is twofold: first, to demonstrate that the SOM-SARIMA-SVR outperforms the rest of possible hybrid or combined models in forecasting the daily number of containers passing through BIP of a maritime port; and second, to test the generalization capabilities of SVR in forecasting logistic tasks, especially the container inspection process that can be a bottleneck in the supply chain. To train the new model, a three-step procedure was developed. In the first step a SOM algorithm allows to divide the database into different clusters or regions. On a second step, a SARIMA model is fitted to the data of each cluster in order to capture the seasonality and the linear behavior. Finally, a SVR model is applied over each obtained cluster considering several hybrid approaches (different input configurations) which can consider the original database and the outputs of the first step (residual and predicted values).

The rest of the paper is organized as follows: The second section gives an overview of self-organizing maps (SOM), seasonal autoregressive integrated moving average (SARIMA) and support vector machines for regression (SVR). The empirical data, the combined-hybrid methodology, and the experimental procedure are presented in Section 3. The fourth section discusses the results obtained. Finally, the last section summarizes the important conclusions of this work.

2. Methods

The proposed model consists of hybridizing a seasonal autoregressive integrating moving average (SARIMA) model as a linear model, and a non-linear model such as a support vector machine for regression (SVR). These methods are two of the most important forecasting models concerning their respective domains. Additionally, a clustering method, Kohonen self-organizing map (SOM), is combined with this hybrid (SARIMA-SVR) model. The methodology and basic concepts are next described.

2.1. Self-Organizing Maps (SOM)

Within the unsupervised learning field, a SOM is a kind of neural network that has gained special attention during the past two decades. First proposed by Kohonen [8,20], a SOM is a classification technique that groups objects of the systems into regions called clusters. This classification is based on the similarity or nearness of these objects without external factors influencing their performance. In the process, the neurons of the model organize themselves considering only those that play a similar

role, forming a cluster. The representative point of this cluster is called centroid, being the central point, which can be used as the center point of a classifier based on minimum distance.

The topology of a SOM model consists of several neurons distributed into two layers: the input and the output layer. The first one is formed by k neurons. Each neuron corresponds with one input. The output layer, called the competition layer, can consist of different topologies (2-D grid for this case). The pre-processing is performed in this layer. In the process, all the neurons j of the output layer are connected by weights $(w_{i,j})$ with all neurons i of the input layer. A weight vector, $w_j = (w_{j,1}, w_{j,2}, \dots, w_{j,k})$, is thereby associated to each output neuron j of the competitive layer (k is thereby the input vector dimension). The training of a SOM model is based on a competitive process where networks are trained iteratively. Different input vectors $x_i = [x_1, x_2, \dots, x_k]^T$ are presented to the network at each iteration. During the network training, the Euclidean distance between x and all the weight vectors are computed as follows:

$$\|x - w_b\| = \min_j \{\|x(t) - w_j(t)\|\}, \quad j = 1, 2, \dots, l, \quad (1)$$

where l is the number of output neurons. According to Equation (1), w_b is considered the winning neuron, i.e., the neuron that has the weight vector closest to x . In addition, the weight of the winning neuron is updated at time $t + 1$, according to Equation (2). The same occurs with its associated neighbor neurons:

$$w_i(t + 1) = \beta_i \cdot \delta_{b,i}(t) \cdot [x(t) - w_i(t)] + w_i(t) \quad (2)$$

where $\delta_{b,i}(t)$ is the neighborhood function (usually a Gaussian function) associated to the neuron i , and $\beta(t)$ is the exponential decay learning factor. Both parameters, $\beta(t)$ and $\delta_{b,i}(t)$, decay with time. The training algorithm stops when the maximum number of epochs (a representation of all inputs patterns) is achieved, or when the performance is minimized to the target.

2.2. Auto-Regressive Integrated Moving Averages (ARIMA)

ARIMA models were introduced by Box and Jenkins [2]. ARIMA has been a widely used forecasting linear model during several decades. In these kind of models, the future value is accepted to be a linear function of several past observation and an error term. Three prediction terms compose this linear function: the autoregressive term (AR), the moving average term (MA), and the integration term (I). A SARIMA model can be obtained by extending the ARIMA model to include seasonal features. In this way, the model is specified as SARIMA(p, d, q)(P, D, Q) $_S$, where q represent the order of the moving average terms, p denotes the order of the autoregressive terms, and d is the degree of differencing. (P, D, Q) deals with the seasonal part and the capital letters corresponds to their counterparts for the seasonal models with the seasonal orders and the seasonality of the model is represented by the parameter s . Equation (3) depicts a typical expression of the SARIMA model:

$$\varphi_p(L)\Phi_P(B^s)\nabla^d\nabla_s^D y_t = \theta_q(B)\Theta_Q(B^s)a_t, \quad (3)$$

where y_t is the observed value, ∇^d and ∇_s^D are the regular and seasonal differencing operators, respectively, p and P are the number of non-seasonal and seasonal autoregressive terms, q and Q are the number of non-seasonal and seasonal moving average terms, d and D are the number of regular and seasonal differences, φ and Φ depict the value weights of the non-seasonal and seasonal autoregressive term, θ and Θ represent the weights of the non-seasonal and seasonal moving average term, the seasonality is represented by S , and a_t is the noise term. The first step is to identify the SARIMA structure assisted by the observation of the simple and partial autocorrelation function (ACF and PACF) of the time series. In addition, data must be stationarity in mean, using power transformations, and stationarity in variance, employing differencing of the time series. Second, the parameters of the model are estimated. Third, the estimated residuals must be checked. The requirements of a white noise process should be satisfied by the residuals and they are assumed to be independent. Several

statistic tests and plots of the residuals are used for this purposes. Finally, an estimation of the future value of the volume of containers is obtained. In this work, a value range of the model parameters are iteratively tested to identify the most suitable model.

2.3. Support Vector Machines for Regression (SVR) Models

In contrast to ANN models, support vector machines (SVM) is a kind of machine learning technique focused on the structural risk minimization instead of the empirical risk minimization principle. The main objective of this method is maximizing the margin distance [21]. The model can be formulated as the following equation:

$$y(x) = w^T \phi(x) + b, \quad (4)$$

where b denotes the bias term and w is the vector of weights. $\phi(x)$ represents the kernel function used to deal with the nonlinear problem, mapping the input data into a higher dimensional (feature) space where data can be linear. First introduced for classification problems, the ε -insensitive loss function, presented in Equation (5), has enabled its use in regression problems:

$$L_\varepsilon(y) \begin{cases} 0, & \text{if } |f(x) - y| \leq \varepsilon \\ |f(x) - y| - \varepsilon, & \text{otherwise} \end{cases}, \quad (5)$$

In Equation (5), ε denotes the area of ε -insensitive. The process is the following: first, the input data are mapped into a new space of higher dimensional features, called feature space, by a non-linear mapping a priori using a kernel transformation. The aim of this feature space is to detect a linear regression function that can be fit the output data with the input data. This linear regression corresponds to the nonlinear regression model in the original space. Equation (6) represents the problem that should be optimized:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i^+ + \xi_i^-) \\ \text{subject to :} \quad & w \cdot x_i + b - y_i \leq \varepsilon + \xi_i^+ \\ & y_i - w \cdot x_i - b \leq \varepsilon + \xi_i^- \\ & \xi_i^+, \xi_i^- \geq 0 \end{aligned} \quad (6)$$

with $i = 1, \dots, l$. ξ_i^- and ξ_i^+ are the slack variables that deal with the training error on the top and the bottom, respectively. These variables take zero values within the $\{-\varepsilon, \varepsilon\}$ area, and non-zero values outside it. $\frac{1}{2} \|w\|^2$ is the structure risk concerning the flatness of the model and the parameter C is a correction factor that deals with the trade-off between the flatness and the error. In this work, the Gaussian kernel was chosen as kernel function. The dual optimization problem can be solved with the Lagrangian multiplier method [21]. Finally, the method obtains the support vectors as the final decision, which are the observations with non-zero coefficients of the Lagrangian multipliers.

3. Forecasting Approach

In this study, a three-step procedure based on a SOM-SARIMA-SVR model is proposed in order to predict the daily number of containers passing through a sanitary facility of a seaport. A further objective is to compare the prediction results of the proposed three-step procedure with the obtained using other possible methodologies, such as the SVR models in a single way, the hybrid SARIMA-SVR model and the combined SOM-SVR model. The proposed approach combines the capability of SARIMA models in capturing the linear patterns and the ability of SVR in modelling nonlinear patterns, together with the advantages of the SOM clustering algorithm.

The experimental database comes from the BIP of the Port of Algeciras Bay, located in the South of Spain. The BIP of Algeciras is the approved facility where import goods are inspected and checked

before entering the Community territory of the European Union. The Port of Algeciras Bay was the first port in the Mediterranean Sea and the fourth port in the European continent related to the total throughput in 2019 (109.4 million tons) only surpassed by the ports of Rotterdam (The Netherlands), Antwerp (Belgium) and Hamburg (Germany). The database was provided by the Port Authority and contains daily records of the number of containers at the Algeciras BIP from 2010 to 2014.

3.1. The Proposed Hybrid Methodology

A SOM algorithm (step I) was firstly used to divide the whole database into several disjoint clusters with similar statistical properties. Then, the seasonality and the linear patterns were captured using a SARIMA model in each cluster (step II). As a result, a set of predicted values and residuals from each cluster was obtained. Finally, a SVR model was then implemented in the third step over each cluster to generalize the non-linear relationship, obtaining the final prediction (step III). Inputs of the SVR model were the original and predicted data from the second (SARIMA) step. Two hybrid approaches were tested in the last step considering the configuration and the number of inputs.

The forecasting performance was assessed using different prediction horizons: for the SARIMA step, only one-day ($ph = 1$) prediction horizon was considered; and for the final (SVR) step, two prediction horizons were tested, one-day ($ph = 1$) and seven-day ($ph = 7$) ahead. The prediction is one-step ahead (y_{t+ph}) in both cases. In the $ph = 1$ case, the prediction is y_{t+1} , and for the $ph = 7$ case the prediction is y_{t+7} . The estimation can be thereby modelled as a nonlinear function of the n preceding values of the time series and an error term. This is called the autoregressive window (n) and its design is presented in Figure 1. The autoregressive window for $ph = 1$ (steps II and III) is presented on the top of the figure and for $ph = 7$ (step III) is showed on the bottom of the figure.

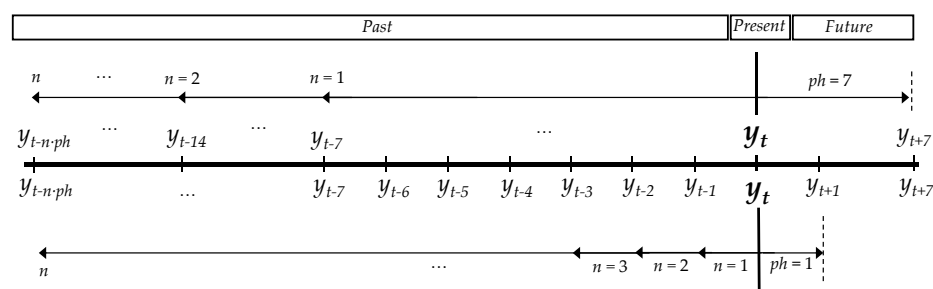


Figure 1. Autoregressive window sizes in Steps II and III and their prediction horizons (ph): one-day prediction horizon (below the timeline) and seven-day prediction horizon (above the timeline). n is the size of the autoregressive window.

For the $ph = 7$ case, the autoregressive window is composed of values of the container series periodically sampled every seven days in the past. This is due to the weekly seasonality found in the analysis of the autocorrelation function of the time series. The main assumption here is that best prediction is obtained when using as inputs several samples from the past of the same day of the week (i.e., using several successive Mondays in the past to predict a future Monday). The proposed SOM-SARIMA-SVR procedure is shown graphically in Figure 2.

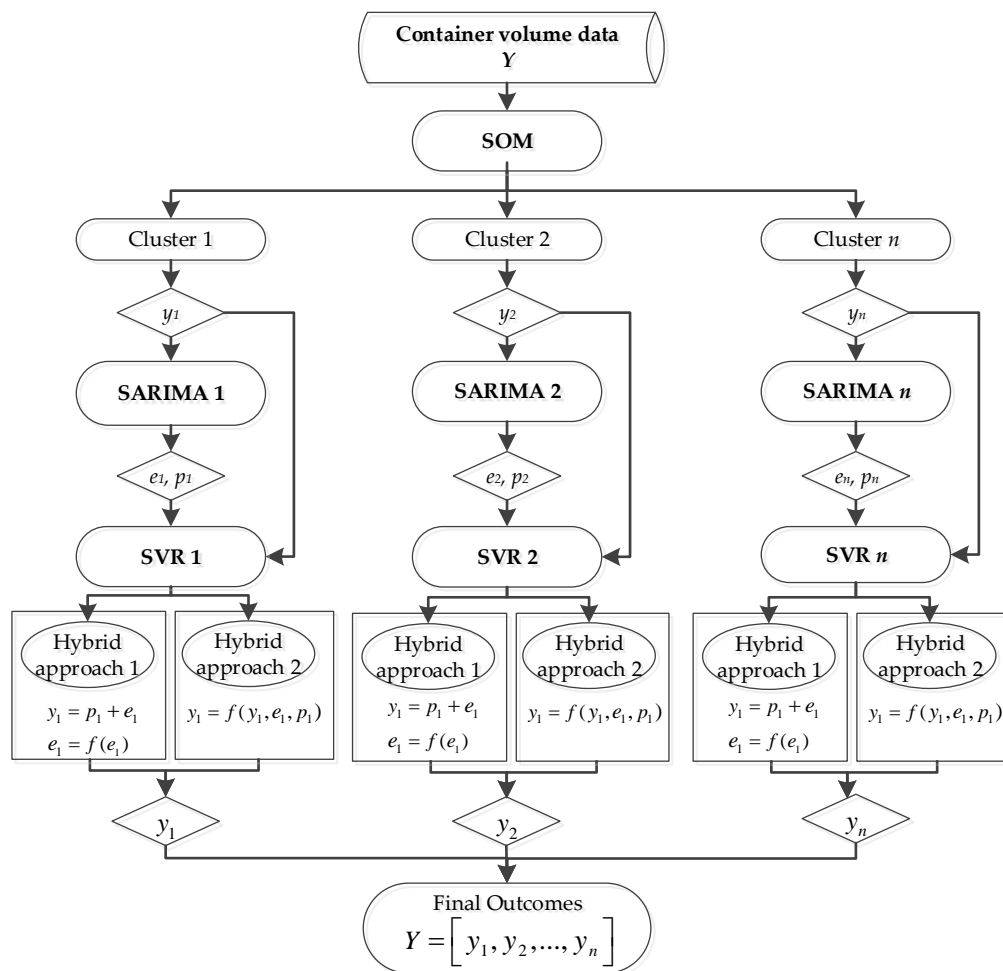


Figure 2. The overall process scheme of the SOM-SARIMA-SVR approach.

3.1.1. Step I: SOM

A self-organizing feature map (SOM) is first applied to the data in order to split the database into several disjoint groups, called clusters, with similar statistical distribution. The assumption is that, in the next step, a forecasting technique can predict more accurately each cluster instead of the whole database. Each cluster works independently in the second and third step. In such cases, a single SARIMA and SVR models are applied independently after decomposing the heterogeneous data into smaller homogeneous regions. A predefined number of clusters, c , are selected, before starting the SOM algorithm, to assess and compare the performance of different solutions.

An experimental framework was developed in order to select the optimal number of past values (nc) of the input vector (the number of input neurons). Thus, a database of inputs was designed as in Equation (7):

$$x_i = [y_t, y_{t-1:k}, y_{t-2:k}, \dots, y_{t-nc:ph}]^T, \quad (7)$$

where t is each sample in the daily time series from 2010 to 2014 and k means the sample period explained graphically in Figure 1 ($k = 1$ or $k = 7$ days). Each row is arranged recursively using different lagged terms (as in an autoregressive window).

A large number of different experiments were performed modifying the input vector size and the number of iterations. In order to consider the inherent randomness of the process, 20 repetitions of each experiment were carried out.

3.1.2. Step II: SARIMA

A SARIMA model is fitted to each cluster generated by the SOM in Step I, obtaining different predicted and residual values of these clusters. The main reason to select SARIMA as a forecasting model is its ability to capture linear patterns. In order to assess the forecasting performance, and due to the deterministic nature of linear models, a hold-out validation technique was applied during the process. The data (of each cluster) was divided into two groups: the training set containing two-thirds of the dataset, and the test set comprising the rest of the samples. The parameters of the model were adjusted using the training set and the test set was used to validate the model. Different values of the model parameters were tested using a trial-and-error procedure. Only the $ph = 1$ prediction horizon was tested. The values of the parameters tested within each cluster are, for the non-seasonal part: $p = 0, 1, 2, 3, 4$; $d = 0, 1, 2$ and $q = 1, 2, 3, 4$; and for the seasonal part: $s = 2, 5, 7$; $P = 0, 1, 2$; $D = 0, 1, 2$ and $Q = 0, 1, 2, 3$.

3.1.3. Step III: SVR

At this stage, a SVR model is applied to each cluster. The three different adjustable parameters, which governs the SVR model, were determined by an iterative process (trial-and-error). The parameters of the different SVR models tested are shown in Table 1. Note that the optimum number of clusters was two. This issue is further explained below.

Table 1. Parameter ranges used for the SVR models in the third step. The subscript indicates the cluster to which the parameter belongs to (1 or 2).

Parameters	Cluster	Values	
Hybrid configuration:		SOM-SARIMA-SVR-1	SOM-SARIMA- SVR-2
Number of variables:			
Input layer	1	e_1	y_1, p_1, e_1
	2	e_2	y_2, p_2, e_2
Output layer	1	e_1	y_1
	2	e_2	y_2
Autoregressive window size (step: 1):			
Residual (e): ne	1, 2	1:20	none, 1:20
Original data (y): ny	1, 2	-	none, 1:10
Predicted values (p): np	1, 2	-	none, 1:20
Parameters:			
ε (step: 1):	1, 2	$2^{(-12:-2)}$	$2^{(-12:-2)}$
γ (step: 1):	1, 2	$2^{(-12:-2)}$	$2^{(-12:-2)}$
C:	1, 2	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 50, 100:1000 (step: 100)	
Prediction horizons	1, 2	1, 7 days	

For each cluster, the inputs of the SVR model are formed by the clustered data of the original time series, y_i , and their forecasted values, p_i , and residuals, e_i , obtained in the second (SARIMA) step (i denotes the cluster they belong to). The presence or absence of these variables in the inputs determine the hybrid configurations. Each variable is sorted recursively as an autoregressive window. The sizes of the original data, predicted values and residuals from the SARIMA step are denoted as ny , np and ne , respectively.

Due to the reduction of available data in the clusters, a randomized resampling strategy was implemented during the SVR process to ensure the independence of the results. For this work, a twofold cross-validation (2-CV) technique was used. This validation strategy was repeated 20 times and the final value of the prediction performances was the average of these repetitions. The 20 repetitions were implemented for each combination of parameter values of the range shown in Table 1. Once all the range values of the parameters were tested, the parameter combination that achieves the best performance

index values is selected. The total predicted time series is formed by adding the predictions of each clusters. Note that, as in the SARIMA model, the best SVR model may be different on each cluster.

Two hybrid approaches were proposed and assessed. The prediction results were obtained for two prediction horizons, $ph = 1$ and $ph = 7$.

SOM-SARIMA-SVR-1 model (hybrid approach 1). This classical approach considers that the relationship between the linear component L_t and the nonlinear component NL_t of a time series is additive. Therefore, the time series can be decomposed in these two independent terms. Then, a linear forecasting model such as SARIMA can be applied in order to model the linear component and thereby to obtain the predicted values, \hat{p}_t , and the residual, e_t . Subsequently, a SVR model is applied over the residuals in order to fit the nonlinear component NL_t . The main assumption of this traditional approach is that the nonlinear relationship of the time series can be found, if any, in the residuals of the linear model. As a consequence, NL_t is a predicted error and it is a function of the residuals obtained from the linear model:

$$NL_{t+ph} = f(e_t, e_{t-ph}, \dots, e_{t-n-ph}) + \varepsilon_t = \hat{e}_{t+ph}, \quad (8)$$

where \hat{e}_t is the predicted residual, f is the nonlinear function obtained by the SVR model, n is the size of the autoregressive window, ph is the prediction horizon, and ε_t is the error term. Finally, the prediction is obtained by adding the two single components:

$$Y_{t+ph} = \hat{L}_{t+ph} + NL_{t+ph}. \quad (9)$$

Other proposals in the framework of these kinds of hybrid models have emerged in recent years. Particularly, incorporating the predicted values from a previous step and the original data with the residuals as inputs of the non-linear forecasting model has gained special attention. The time series is considered as a function of several variables, such as the original data, the residuals, and the predicted values from the linear model. Some examples of this two-step methodology are the works of Khasei and Bijari [15,21]. Based on these last works, the following hybrid approach 2 was developed (after applying the SOM method) considering two or three types of variables as inputs.

SOM-SARIMA-SVR-2 model (hybrid approach 2). The time series is here considered a nonlinear function of the original data and the residuals and the predicted values from the first step:

$$\hat{Y}_{t+ph} = f(y, e, \hat{p}) + \varepsilon_{t+ph}, \quad (10)$$

where f is the nonlinear function obtained with the SVR model, y_t is the original data, e_t is the residual obtained in the SARIMA model, and \hat{p} is the predicted value from the SARIMA model at time t . These variables are presented in an autoregressive form, as it is expressed in the following equation:

$$\hat{Y}_{t+ph} = f(y_t, y_{t-1-ph}, y_{t-2-ph}, \dots, y_{t-ny-ph}, e_{t+ph}, e_t, e_{t-1-ph}, \dots, e_{t-ne-ph}, \hat{p}_{t+ph}, \hat{p}_t, \hat{p}_{t-1-ph}, \dots, \hat{p}_{t-np-ph}) + \varepsilon_{t+ph} \quad (11)$$

where ne , ny and np represent autoregressive window sizes for e , y and \hat{p} variables, respectively. Considering two or three variables, and their autoregressive window sizes, a large number of possible functions can be tested.

These two hybrid approaches are used in each clusters formed in the second step. Considering two clusters, the parameter ranges used are also presented in Table 1. Note that in cases which variables e and p are involved, their first value used as input is the predicted value at $t + ph$ time. The same cannot be said for the original data variable y , due to the y value at time $t + ph$ is the final prediction pursued. In all instances, the SARIMA model is necessary to obtain the required data used as inputs in the third step, together with the original patterns of the container series.

3.2. Performance Indexes

3.2.1. Performance Criteria of Stage I (Clustering)

The effectiveness of the clustering process was measured using two clustering quality indices (CQI_s) based on the Silhouette function (S_i). The silhouette value determines the degree of equality in clusters between a selected pattern and the other patterns that conform a certain cluster compared with the patterns in other clusters. The S_i value for each pattern is between -1 to $+1$ and is defined in Equation (12):

$$S_i = D_i - d_i / \max(d_i, D_i), \quad (12)$$

where D_i is the minimum average distance from one pattern of a cluster to another pattern in another cluster and d_i is the average distance in the own cluster from one pattern to the rest of patterns. The two CQI_s , which are based on the mean and the sum of the silhouette values of all clusters respectively, are collected in Equations (13) and (14). The lower of these values, the worst of the clustering results:

$$QI_1 = (\overline{S_i}), \quad (13)$$

$$QI_2 = \sum S_i. \quad (14)$$

3.2.2. Performance Indexes of Prediction Stages

In order to compare the forecasting performance of multiple models, the index of agreement (d), the autocorrelation coefficient (R), the mean square error (MSE), the mean absolute percentage error ($MAPE$), and the mean absolute error (MAE) are the performance indices that were considered to calculate the estimation of the generalization error in the prediction stages (steps I and III). Equations (15)–(19) shows these performance criteria and their calculation, where m is the sample size, y_i is the real value of the observation and \hat{y}_i is its corresponding predicted value. R values are comprised between -1 and 1 whereas d values are between 0 and 1 . These indices measure the correlation between predicted and real values. By contrast, MAE , $MAPE$ and MSE measure the deviation between predicted and real value. d and R values close to 0 indicate a worse performance whereas lower MAE , $MAPE$ and MSE values indicate better performance:

$$d = 1 - \frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|)^2} \quad (15)$$

$$R = \frac{\sum_{i=1}^m (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^m (y_i - \bar{y})^2 \cdot \sum_{i=1}^m (\hat{y}_i - \bar{\hat{y}})^2}} \quad (16)$$

$$MSE = 1/m \cdot \sum_{i=1}^m (y_i - \hat{y}_i)^2, \quad (17)$$

$$MAE = 1/m \cdot \sum_{i=1}^m |\hat{y}_i - y_i|, \quad (18)$$

$$MAPE = 1/m \cdot \sum_{i=1}^m |(y_i - \hat{y}_i) / y_i|. \quad (19)$$

4. Experimental Results and Discussion

The results of the proposed hybrid SARIMA-SOM-SVR models to predict the daily number of containers passing through the BIP of the Port of Algeciras Bay are presented and discussed in this section. These results are those obtained in the third step of the proposed procedure and they were assessed and compared with those achieved with the single SVR, the combined SOM-SVR, and the hybrid SARIMA-SVR models. Nevertheless, a cursory discussion of the previous results obtained at the first and the second steps (SARIMA and SOM stages) is given in this section for a general overview. The prediction was one-step ahead and two prediction horizons were considered, $ph = 1$ and $ph = 7$ days.

First, a two-dimensional competitive Kohonen network (SOM 2-D) was employed as a clustering technique. The final aim of this step is to divide the whole database into several disjoint clusters with similar statistical features. The assumption is that the prediction of a smaller cluster can be easier than the prediction of the whole database. Testing different configurations of the SOM network, the most appropriate SOM map size for the data was 8×8 neurons in the output layer with a hexagonal grid topology and a three-dimensional input space.

Figure 3a represents the SOM weight positions, which shows the locations of the data points and the weight vectors after the SOM algorithm was trained. Grey and green points represent neurons and input vectors, respectively. Red lines are the connections between neurons. Although the figure suggests the existence of two different groups (a group appears in the bottom-left region and another more disperse group in the central region), not all the weights can be visualized at the same time (one weight for each element of the input vector). Thus, the SOM neighbor distances could be useful (Figure 3b). This figure indicates the distances between neighboring neurons, where grey hexagons represent the neurons (8×8 map), connecting themselves by the red lines. The color of the segments containing these red lines indicates the distance between neighboring neurons. Lighter colors represent smaller distances, whereas larger distances are represented by darker colors. As the figure indicates, a group of light segments appear in the bottom-right region. This region is bounded by some dark segments. According with the previous figure, this situation indicates that the SOM network has clustered the data into two groups.

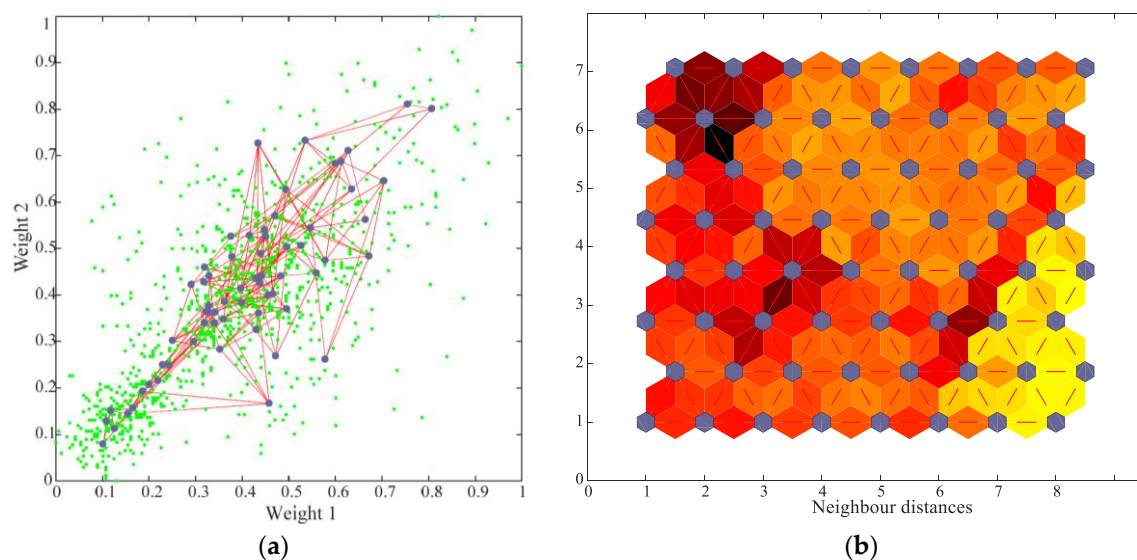


Figure 3. (a) The SOM weight positions and (b) the SOM neighbor distances between neurons.

These results can be contrasted analytically and are collected in Table 2, which shows the best results obtained per cluster and its input-vector configuration. Based on two clustering performance indexes (CQI_1 , and CQI_2), the two-classes clustering was confirmed to be the best choice for the time series, reaching the highest values of CQI_1 and CQI_2 (0.659 and 717.548, respectively). Consequently, the database was divided into two groups, hereinafter called Cluster 1 and Cluster 2. These results were achieved using a three-element input vector ($nc = 3$) with a temporal leap of 7-day in the past. Therefore, the input vector x_t is comprised of the actual value of the time series y_t , and two 7-day lagged past value (y_{t-7} and y_{t-14}), conforming a three-dimensional input space. That is, $x_t = [y_t, y_{t-7}, y_{t-14}]^T$, where $t = 1, 2, \dots, 1461$.

Table 2. Clustering results of the SOM step, where c is the number of clusters tested and nc is the size of the input vector. The temporal leap in the past is 1 or 7 days.

Best Configuration		Performance Indices		
Clusters (c)	Temporal Leap	nc	CQ1	CQ2
2	7	3	0.659	717.548
3	7	3	0.627	682.549
4	1	3	0.588	643.792
5	1	3	0.601	658.248

A SARIMA model was applied independently in each cluster in the second step. Considering the nature of the time series of each cluster, the clustered data was stationarized in mean (using a logarithmic transformation) and variance (using second differencing for Cluster 1 and second seasonal differencing for Cluster 2). Then, using an iterative trial-and-error procedure, the best-fitted models were ARIMA(2,0,3) for Cluster 1 (without seasonal part) and SARIMA(2,1,2)(2,1,3)₅, with a seasonality of 5 days for Cluster 2. Although these linear models obtained the best values of the performance indexes, they have to be validated before continuing to step III. The residuals of the model must be satisfied the requirements of a white noise process. Therefore, the residuals of both clusters were checked. On the one hand, Figure 4 left collects the residual diagnosis plot of Cluster 1. According to Figure 4a left, the residuals have a mean around zero whereas Figure 4b left shows no obvious violation of the normality assumption. Figure 4c,d left show no significant autocorrelation and residuals are thereby uncorrelated. On the other hand, Figure 4 right shows the residual diagnosis plot of Cluster 2, corresponding to SARIMA(2,1,2)(2,1,3)₅ model. As in the previous case, residuals satisfy the requirements of a white noise process. In conclusion, the ARIMA(2,2,3) and SARIMA(2,1,2)(2,1,3)₅ models were validated as linear models to Cluster 1 and 2, respectively, in the second step. Once the ARIMA model is applied, a set of residuals and predicted values are obtained for each cluster. These values, together with the original data, are used as inputs in the third step.

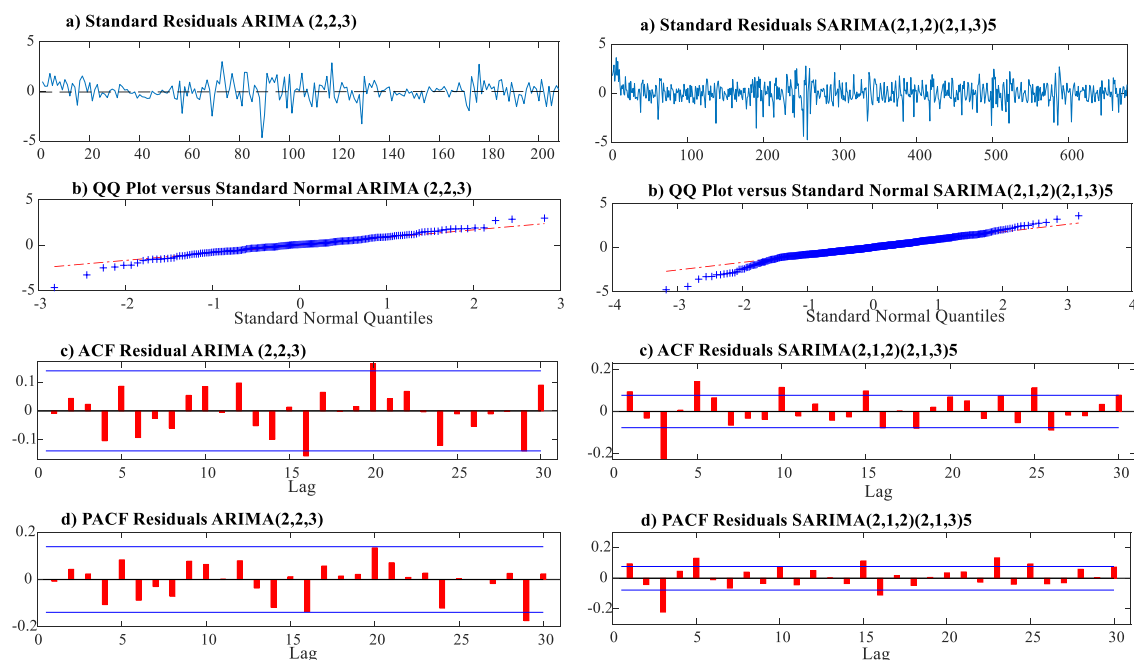


Figure 4. Results of the residual diagnosis for Cluster 1 (left), and for Cluster 2 (right): (a) standard residual plot; (b) quantile–quantile plot of standard residual versus the normal distribution; (c) sample ACF of the residual and (d) partial ACF of the residual.

Finally, in the third step a SVR model was applied to each cluster considering the two proposed hybrid approaches, which were configured depending on the input setting used. Considering each hybrid configuration separately, a most accurate SVR model was achieved in each cluster to fit the data. Two hybrid approaches with two prediction horizons were considered in this third step for each cluster. Therefore, four prediction schemes for each cluster were evaluated (two hybrid approaches for each cluster and two prediction horizons for each hybrid approach). The final prediction results of each hybrid approach were obtained by joining the prediction values achieved in the two clusters as a single predicted time series. That is:

$$\hat{y}_{inspections} = \{\hat{y}_{Cluster1}\} \cup \{\hat{y}_{Cluster2}\}, \quad (20)$$

The most accurate models for each hybrid approach are collected in Table 3. The prediction results in terms of performance indexes reflect those obtained once the predicted values of the two clusters were grouped. Table 3 is divided according to the prediction horizon used ($ph = 1$ or $ph = 7$ days). Additionally, for each prediction horizon, results are collected depending on the hybrid configuration applied. For the hybrid approach 2 (SOM-SARIMA-SVR-2), results are presented considering the inputs employed in the model (y , e or p) in order to demonstrate which inputs are most relevant. Regardless of the parameter configuration, only the best value achieved in each performance index is depicted in both prediction horizons for a better understanding. The best overall value of each performance index and the best hybrid configuration are pointed out in bold.

Table 3. Mean prediction performance results of the proposed SOM-SARIMA-SVR models for one-day and seven-day prediction horizons. The final number of the model indicates the hybrid approach (1 or 2). The column inputs indicate the inputs used in the SVR models, where y , e , and p represent the original data, residuals and predicted values from the second step, respectively. Best values in bold.

ph	Hybrid Approach	Model	Inputs	Performance Indices				
				R	d	MSE	MAE	MAPE
1	1	SOM-SARIMA-SVR 1	e	0.8039	0.8917	296.3551	11.8413	18.6794
1	2	SOM-SARIMA-SVR 2	y, e	0.8101	0.8936	287.0105	11.6790	17.9391
1	2	SOM-SARIMA-SVR 2	y, p	0.8024	0.8914	288.7795	11.7174	17.7721
1	2	SOM-SARIMA-SVR 2	y, e, p	0.8054	0.8919	287.8423	11.8362	17.8603
7	1	SOM-SARIMA-SVR 1	e	0.8037	0.8888	299.6534	11.9761	18.4053
7	2	SOM-SARIMA-SVR 2	y, e	0.8051	0.8912	290.8783	11.8295	18.1178
7	2	SOM-SARIMA-SVR 2	y, p	0.7906	0.8803	306.1622	12.4121	18.3690
7	2	SOM-SARIMA-SVR 2	y, e, p	0.8101	0.8920	289.7224	11.8394	17.9783

The SOM-SARIMA-SVR-2 (the one without p as input) provides the most accurate results for one-day ahead prediction, followed by the rest of possible models of the hybrid approach 2 (considering different inputs) and finally the hybrid approach 1, in that order. This hybrid approach achieved the best value in at least four performance indexes. In this case, more sophisticated models obtained no better results. Nevertheless, the classical approach considers an additive relationship between the linear and nonlinear component of the time series. Consequently, this is less powerful than the other approach. For one-day ahead predictions, two different input variables (y and e) are proved sufficient to predict the time series accurately. However, there are not great differences among the prediction performances with the rest models.

Similar results were obtained considering the behavior of the models for 7-day ahead prediction, where better values of performance indexes were reached with the hybrid approach 2. The most complex approach (SOM-SARIMA-SVR 2 with all variables as inputs) obtained the best results, reaching four of the five best performance indexes. Surprisingly, the classical approach SARIMA-SOM-SVR-1 is positioned as the third best model, obtaining a great MAE value. However, the results from this

prediction horizon do not outperform those obtained using one-day prediction horizon though the difference is not very significant.

Although all the best performance indexes were gained with one-day ahead predictions, good forecasting performance were also obtained considering both prediction horizons and there exist no major differences among the hybrid approaches. Moreover, the use of a certain set of input variables from the second step (SARIMA set) does not seem to be very relevant to improve the performance of the models in the third step. This could be due to the clustering procedure of the second step, which grouped the variables with similar properties, minimizing the negative or positive impact of selecting a certain set of input variables. Nevertheless, better results in any situation were yielded using the more sophisticated models (hybrid approach 2) instead of the classical approach (hybrid approach 1). Particularly, SARIMA-SOM-SVR-2 with variables e and y as inputs of the SVR achieved the best results.

It is worth mentioning that each performance index of any hybrid approach were computed using different configurations of the parameters. That is, for instance, with the SOM-SARIMA-SVR-2 model (with e and y as input variables), the configuration of the SVR parameters (C , ε and γ) that reached the best value of d is not the same as the one obtained the better value of MAE. It may be caused by the high number of model tested, approximately about 10^6 models for each prediction scheme, where there are parameter configurations almost equal. In this sense and in order to get the smallest MAE value (11.679), the best-fitted network of Cluster 1 for this hybrid configuration 2 in the third step is composed by autoregressive window sizes of twelve for the y input variable ($ny = 12$) and two for the e input from SARIMA step ($ne = 2$), being the optimal SVR parameters $C = 200$, $\gamma = 2^{-4}$ and $\varepsilon = 2^{-8}$. To model Cluster 2, the best parameter configuration was $ny = 12$, $ne = 2$, $C = 50$, $\gamma = 2^{-2}$ and $\varepsilon = 2^{-2}$. For this network architecture, the number and size of SVR inputs coincide in both clusters. The final prediction is reached joining the predicted values of the two clusters. The best network architecture of the hybrid approach 2 is plotted in Figure 5.

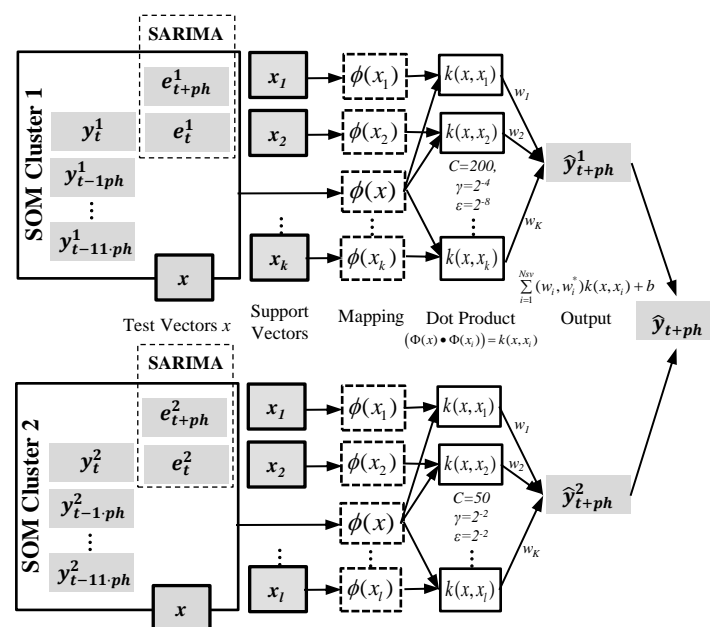


Figure 5. Best-fitting SVR network architecture for each cluster of SOM-SARIMA-SVR-2 hybrid model in the third step ($ph = 1$ case). C , γ and ε are the SVR parameters. Only the inputs y and e are used.

To conclude, a complete comparison among other hybrid models proposed in literature was performed in order to gain a comprehensive view of the real improvement established by the proposed model. Then, the most accurate single SVR model, the most accurate combined SOM-SVR model and the most accurate hybrid SARIMA-SVR model were also compared against the most accurate proposed model. These comparisons are summarized in Table 4, where values and models in bold indicate

the best overall performance index and model, respectively, considering both prediction horizons. As Table 4 shows, the proposed SOM-SARIMA-SVR model outperforms the rest of the models in both prediction horizons. The SOM-SARIMA-SVR model is more competitive than the other models in four of the five performance indexes. Interestingly, another model achieves the highest value in one of the five performance index (SARIMA-SVR for d and $ph = 1$; SOM-SVR for R and $ph = 7$). It can be concluded that the proposed model performs significantly better than the rest of models, especially in terms of MSE, MAE and MAPE values. This suggest that the “divide-and-conquer” principle, introduced with the usage of the clustering stage, can improve the performance of the hybrid models that consider the hybridization of linear and nonlinear forecasting techniques. The proposed methodology combines the strengths of the combined forecasting models with a clustering stage (clustering + forecasting technique) with the advantages of hybrid forecasting models (linear + nonlinear forecasting technique). The combined-hybrid approach has proved to be more effective than the other models in forecasting tasks and particularly in predicting the daily number of containers to be inspected at BIPs in ports and overcomes SOM-SVR model.

Table 4. Comparison of the best mean prediction performance results of the single models (SVR), the combined models (SOM-SVR), the hybrid model (SARIMA-SVR) and the proposed model (SOM-SARIMA-SVR) for one-day and seven-day prediction horizon. Best values in bold.

ph	Model	Performance Indices				
		R	d	MSE	MAE	MAPE
1	SVR	0.7989	0.8836	389.1624	14.3328	23.0695
1	SOM-SVR	0.8042	0.8862	381.6965	14.1565	21.8681
1	SARIMA-SVR	0.8048	0.8952	302.0054	12.0024	19.4246
1	SOM-SARIMA-SVR	0.8101	0.8936	287.0105	11.6790	17.7721
7	SVR	0.8041	0.8880	365.4109	13.7256	21.9900
7	SOM-SVR	0.8120	0.8917	359.7340	13.5181	21.3260
7	SARIMA-SVR	0.7969	0.8872	306.1086	12.0327	19.7383
7	SOM-SARIMA-SVR	0.8101	0.8927	289.7224	11.8295	17.9783

Figure 6 represents a comparison point-to-point between the observed and predicted values for the best-fitted models concerning the $ph = 1$ (at the top) and $ph = 7$ case (at the bottom), respectively. An example period of five weeks of the year 2012 is represented. These figures clearly show that the proposed model produces a better fit to the data. The SARIMA-SOM-SVR model provides a better prediction of the daily number of containers passing through the BIP even in cases where the values are more difficult to predict.

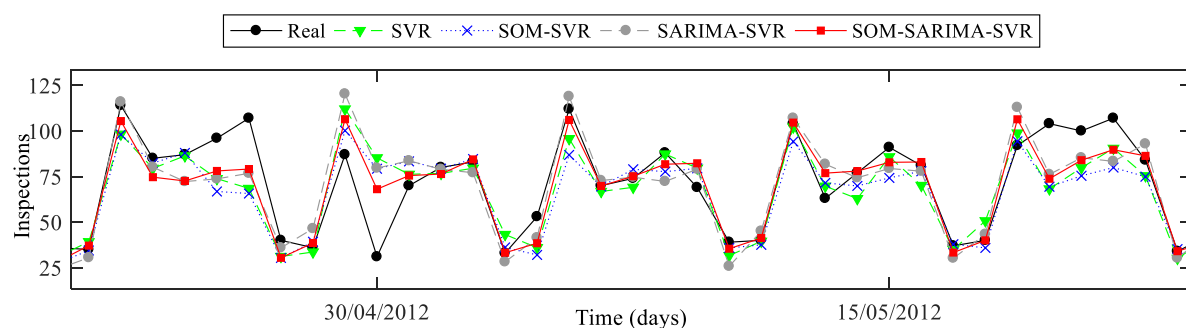


Figure 6. Comparison of the observed and predicted value with the most accurate models. $ph = 1$ case.

5. Conclusions

Prediction tasks and forecasting methodologies and models are constantly evolving. Diverse methodologies to address forecasting problems have been proposed with mixed results, highlighting

the combination of models and the hybridization of them. In this study, a combined-hybrid SOM-SARIMA-SVR forecasting model has been proposed based on a three-step procedure to predict the daily number of containers passing through a Border Inspection Post of a maritime port.

To reduce the complexity of the problem, a clustering SOM is first applied to obtain smaller regions with similar statistical features which may be easier to predict. A SARIMA model is then fitted within each cluster to obtain predicted values and residuals of the clustered database. Finally, a SVR model is used to forecast each cluster independently using the variables obtained from the second step together with the original data as inputs. The combination of each cluster results in the whole predicted time series. The above methodology involves the advantages of combining a forecasting model with a clustering technique and the strengths of the hybrid models in capture linear and nonlinear patterns.

The proposed SOM-SARIMA-SVR model has been developed and compared to other possible methodologies implied in the process (SVR, SOM-SVR and SARIMA-SVR). The results showed that the SOM-SARIMA-SVR model was the most competitive model, improving the forecasting performance of the rest of the models concerning the prediction of the container demand, outperforming these methodologies. Particularly, considering the SOM-SARIMA-SVR model, two hybrid approaches were assessed: the classical additive approach, where the error term (e) of the linear model is the input of the nonlinear model; and the proposed approach, where the prediction is a function of the original data (y) and the error term (e), and the prediction of the linear model (p). Most accurate results were yielded by the second approach in all cases tested. In addition, there are no significant differences between the prediction performance using both prediction horizons. The same occurs with the introduction of certain variables as inputs. These outcomes highlight the robustness of the model. This investigation suggests thereby that the proposed hybrid approach achieves better outcomes getting higher forecasting performance than the classical additive hybrid approach.

To conclude, this study is the first one in using the SOM-SARIMA-SVR model to forecast the volume of containers passing through a BIP of ports in particular, and for time series forecasting in general with promising results. Due to its ability to seize the strengths of linear and nonlinear models and thereby to capture linear and nonlinear patterns, the proposed model could be applied in other time series where these patterns are jointly presented. Particularly, the use of a clustering technique allows reducing the complexity of the time series, increasing the accuracy of the final prediction. Obviously, some limitations are found in the model. As with any data-driven model, when completely different inputs come into the model, the prediction accuracy could worse. Thus, a retrained model and a readjustment of the parameters might be required over time.

Future works will focus on the application of other latest non-linear techniques, such as Deep Learning, in order to assess any improvements in the prediction. Knowing the daily container demand in advance allows detecting workload peaks in a port facility. This guarantees the correct planning and organization of available human and material resources. The proposed methodology can provide an automatic tool to predict workloads at sanitary facilities avoiding congestion and delays. Therefore, it can be used as a decision-making tool by port managers due to its capacity to plan resources in advance.

Author Contributions: Conceptualization, J.J.R.-A. and I.T.; methodology, J.J.R.-A.; software, J.J.R.-A., J.A.M.-L., D.U.; validation, J.J.R.-A., D.U. and J.G.-E.; formal analysis, J.J.R.-A.; investigation, J.J.R.-A. and J.A.M.-L.; resources, D.U. and J.G.-E.; data curation, J.J.R.-A. and J.A.M.-L.; writing—original draft preparation, J.J.R.-A. and J.A.M.-L.; writing—review and editing, D.U., J.G.-E. and I.T.; visualization, J.J.R.-A.; supervision, J.J.R.-A. and I.T.; project administration, I.T.; funding acquisition, I.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by MICINN (Ministerio de Ciencia e Innovación-Spain), grant number RTI2018-098160-B-I00.

Acknowledgments: The database was kindly provided by the Port of Algeciras Bay Authority.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xie, G.; Wang, S.; Zhao, Y.; Lai, K.K. Hybrid approaches based on LSSVR model for container throughput forecasting: A comparative study. *Appl. Soft Comput.* **2013**, *13*, 2232–2241. [\[CrossRef\]](#)
2. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Holden-Day: Oakland, CA, USA, 1976.
3. Gambardella, L.M.; Bontempi, G.; Taillard, E.; Romanengo, D.; Raso, G.; Piermari, P. Simulation and forecasting in intermodal container terminal. In *Proceedings of the 8th European Simulation Symposium*; SCS International: Genoa, Italy, 1976; pp. 626–630.
4. Babcock, M.W.; Lu, X. Forecasting inland waterway grain traffic. *Transp. Res. Part E Logist. Transp. Rev.* **2002**, *38*, 65–74. [\[CrossRef\]](#)
5. Klein, A. Forecasting the Antwerp maritime traffic flows using transformations and intervention models. *J. Forecast.* **1998**, *15*, 395–412. [\[CrossRef\]](#)
6. Peng, W.-Y.; Chu, C.-W. A comparison of univariate methods for forecasting container throughput volumes. *Math. Comput. Model.* **2009**, *50*, 1045–1057. [\[CrossRef\]](#)
7. Zhang, G.P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175. [\[CrossRef\]](#)
8. Kohonen, T. *Self Organising Maps*; Springer: Berlin, Germany, 1995.
9. De Bodt, E.; Cottrell, M.; Letremy, P.; Verleysen, M. On the use of self-organizing maps to accelerate vector quantization. *Neurocomputing* **2004**, *56*, 187–203. [\[CrossRef\]](#)
10. Vlahogianni, E.I.; Karlaftis, M.G.; Golias, J.C. Temporal Evolution of Short-Term Urban Traffic Flow: A Nonlinear Dynamics Approach. *Comput. Civ. Infrastruct. Eng.* **2008**, *23*, 536–548. [\[CrossRef\]](#)
11. Chen, H.; Grant-Muller, S.; Mussone, L.; Montgomery, F. A study of hybrid neural network approaches and the effects of missing data on traffic forecasting. *Neural Comput. Appl.* **2001**, *10*, 277–286. [\[CrossRef\]](#)
12. Ismail, S.; Shabri, A.; Samsudin, R. A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting. *Expert Syst. Appl.* **2011**, *38*, 10574–10578. [\[CrossRef\]](#)
13. Ma, Z.; Xing, J.; Mesbah, M.; Ferreira, L. Predicting short-term bus passenger demand using a pattern hybrid approach. *Transp. Res. Part C Emerg. Technol.* **2014**, *39*, 148–163. [\[CrossRef\]](#)
14. Tseng, F.-M.; Yu, H.-C.; Tzeng, G.-H. Combining neural network model with seasonal time series ARIMA model. *Technol. Forecast. Soc. Chang.* **2002**, *69*, 71–87. [\[CrossRef\]](#)
15. Khashei, M.; Bijari, M. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Appl. Soft Comput.* **2011**, *11*, 2664–2675. [\[CrossRef\]](#)
16. Wang, J.-J.; Wang, J.-Z.; Zhang, Z.-G.; Guo, S.-P. Stock index forecasting based on a hybrid model. *Omega* **2012**, *40*, 758–766. [\[CrossRef\]](#)
17. Chen, K.-Y.; Wang, C.-H. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Syst. Appl.* **2007**, *32*, 254–264. [\[CrossRef\]](#)
18. Zhu, B.; Wei, Y. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega* **2013**, *41*, 517–524. [\[CrossRef\]](#)
19. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69. [\[CrossRef\]](#)
20. Vapnik, V.N. *Statistical Learning Theory*; John Wiley and Sons: New York, NY, USA, 1998.
21. Khashei, M.; Bijari, M. An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Syst. Appl.* **2010**, *37*, 479–489. [\[CrossRef\]](#)

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).